DOCUMENT RESUME

ED 475 836 TM 034 861

AUTHOR	Newman, Mark; Ambrose, Kate; Corner, Trevor; Evans, Jeff; Morris-Vincent, Phyl; Quinn, Shelia; Stephenson, John; Vernon, Lesley
TITLE	Evaluating Educational Impact: The Approach Followed by the Project on the Effectiveness of Problem Based Learning (PEPBL).
INSTITUTION	Middlesex Univ., London (England).
SPONS AGENCY	Economic and Social Research Council, Lancaster (England).
PUB DATE	2003-04-00
NOTE	25p.; Paper presented at the Annual Meeting of the American Educational Research Council (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Definitions; Educational Philosophy; Field Studies; *Instructional Effectiveness; *Outcomes of Education; *Problem Based Learning; Teaching Methods
IDENTIFIERS	England

ABSTRACT

Problem Based Learning (PBL) has been adopted around the world as a philosophy and method for teaching and learning in professional education in particular. Advocates of the approach have made many claims for its success. Despite the apparent widespread use of this approach and the plethora of published papers on PBL, there are numerous basic questions about the method that remain controversial. At a fundamental level there is no universal agreement about what PBL actually is. There is little agreement about what the specific measurable outcomes of PBL are or how they should be measured. These conceptual, methodological, and practical problems were tackled in the Project on the Effectiveness of Problem Based Learning (PEPBL). This paper explains the rationale for the field trial and outlines the research design and methods used in the study as an illustration of one approach to the issue of assessing impact on student learning. (Contains 90 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 475 836

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
M. Newman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Evaluating educational impact: The approach followed by The Project on the Effectiveness of Problem Based Learning (PEPBL)

Mark Newman[1,2], Kate Ambrose[1], Trevor Corner[2], Jeff Evans[3], Phyl Morris-Vincent[1], Sheila Quinn[1], John Stephenson[4], Lesley Vernon[1].

1. School of Health and Social Sciences, Middlesex University
2. School of Lifelong Learning & Education Middlesex University
3. Middlesex University Business School
4. International Centre for Learner Managed Learning, Middlesex University

Paper prepared for the American Educational Research Association Conference
April 2003 Chicago, USA

Copies of the paper can downloaded from:

PEPBL Home page
http://www.hebes.mdx.ac.uk/teaching/Research/PEPBL/index.htm

International Centre for Learner Managed Learning at Middlesex University
http://www.iclml.com/

This study was funded by the ESRC Teaching & learning Research Programme
http://www.tlrp.org/

TM034861

MIDDLESEX UNIVERSITY

T·L·R·P Teaching&Learning Research Programme

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Evaluating educational impact: The approach followed by The Project on the Effectiveness of Problem Based Learning (PEPBL)

Mark Newman[1,2], Kate Ambrose[1], Trevor Corner[2], Jeff Evans[3], Phyl Morris-Vincent[1], Sheila Quinn[1], John Stephenson[4], Lesley Vernon[1].

2. School of Health and Social Sciences, Middlesex University
2. School of Lifelong Learning & Education Middlesex University
4. Middlesex University Business School
5. International Centre for Learner Managed Learning, Middlesex University

Keywords
Study design, measurement, research methods, learning outcomes,

Abstract

Problem Based Learning (PBL) has been adopted around the world as a philosophy and method for teaching and learning in professional education in particular. Advocates of the approach have made many claims for its success. Despite the apparent widespread use of this approach and the plethora of published papers on PBL there are numerous basic questions about the method that remain controversial. At a fundamental level there is no universal agreement about what PBL actually is. Similarly there is little agreement about what the specific measurable outcomes of PBL are or how they should be measured. These conceptual, methodological and practical problems were tackled in the Project on the Effectiveness of Problem Based Learning (PEPBL), funded by the ESRC's Teaching and Learning Research Programme. This paper explains the rational for the field trial and outlines the research design and methods used in the study as an illustration of one approach to the issue of 'Assessing Impact on Student Learning' being used in the TLRP funded programmes.

Introduction

Problem-Based Learning (PBL) provides an alternative philosophy and method for and has been introduced into education in many professional fields including medicine, nursing, dentistry, social work, management, engineering and architecture. In its modern guise PBL started to become a feature of educational programmes during the 1960's. Since then there has been a steady growth in the number of programmes and institutions that have adopted PBL around the world. This transformation has been encouraged by an almost evangelical PBL movement that has published of a wealth of anecdotal material extolling the virtues of PBL (Wilkie 2000). PBL has been endorsed by a wide variety of national and international organizations (Tompkins 2001). These include the Association American Medical Colleges (Muller 1984) the World Federation of Medical Education (Walton & Matthews 1989), The World Health Organization (World Health Organization 1993), the World Bank {1993} and the English National Board for Nursing Midwifery and Health Visiting (English National Board 1994). In recent years the advantages that are claimed for PBL have become part of the generally articulated outcomes for education at all levels (Evenson & Hmelo 2000)

The theoretical basis of PBL

The philosophical and theoretical underpinnings of PBL were not explicit in the early PBL literature (Rideout 2001). Barrows, a pioneer of PBL, explains that the he and the other developers of the original the McMaster PBL curriculum had no background in educational psychology or cognitive science. They just thought that learning in small groups through the use of clinical problems would make medical education more interesting and relevant for their students (Barrows 2000). PBL can be interpreted as congruent with at least two distinct streams of theory about knowledge and learning, Constructivism (Evenson & Hmelo 2000) and Cognitive Psychology (Schmidt 1993).

PBL

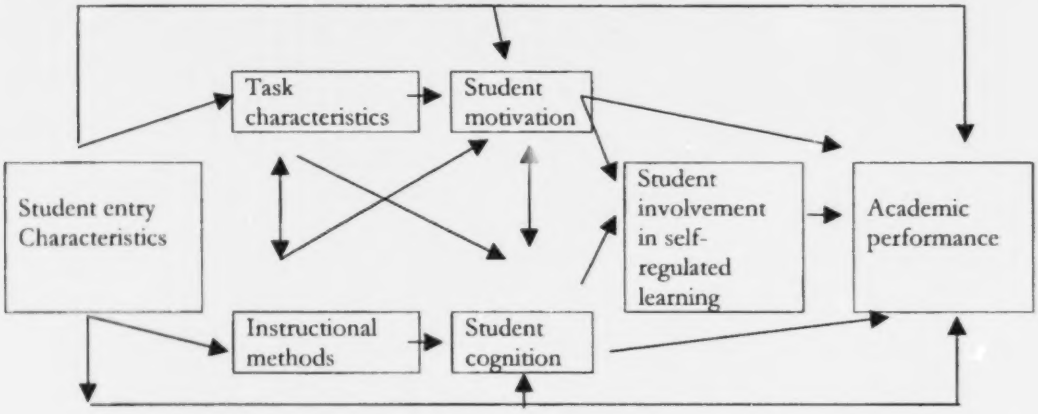
The wide dissemination of PBL has spawned many variations (Barrows 2000). In a review of the field Vernon and Blake (1993), found that PBL was described in a variety of ways that could be summarised as a complex mixture of general teaching philosophy, learning objectives and goals and faculty attitudes and values. Maudsley (1999) argues that the label PBL is often borrowed for prestige or subversion, adorning many narrowly focused single subject courses within traditional curricula that do not use PBL at all. This would seem to be supported by the findings of a review of the curricula of American Medical Schools that claimed to use PBL which found that PBL was being used as a generic category which included almost any teaching approach (Myers Kelson & Distlehorst 2000).

Bereiter and Scardamalia (2000) distinguish between PBL (uppercase) and pbl (lowercase). Lowercase pbl refers to an indefinite range of educational approaches that give problems a central place in the learning activity. Practitioners of PBL tend to adhere to the structures and procedures systematized by Barrows (Barrows 1986). Engel (1991), Barrows {1986} and Savin-Baden (2000) all emphasize that the difference with PBL is at the level of curriculum. Walton and Matthews (1989) argue that PBL is to be understood as a general educational strategy rather than merely a teaching approach. They present three broad areas of differentiation between PBL and the 'traditional' subject centred approaches:

1. Curricula organization: Around problems rather than disciplines, integrated, emphasis on cognitive skills as well as knowledge.
2. Learning environment: use of small groups, tutorial instruction, active learning, student centred, independent study, use of relevant 'problems'.
3. Outcomes: Focus on skills development and motivation, abilities for life long learning

Viewed in this way PBL can be conceptualised as a carefully designed system of teaching and learning selected to support particular types of learning through attention to factors that have been identified as affecting academic performance (see figure 1) (Entwistle 1992).

Figure 1: General model of college teaching and learning (Mckeachie et al. 1986)



PBL in the classroom

The curriculum is operationalized into a number of scenarios or problems. The scenarios are designed to mirror situations that the students will encounter in 'real life'. In addition to a short narrative a scenario pack typically includes additional information pertinent to 'the case' and a directory of further resources (see box 1). The scenarios provide the triggers for the students together with their tutor to embark on the process of learning. The tutor maybe given a list of learning issues that the scenario can be used to generate.

The teaching and learning process used in PBL is described by various authors in terms of a number of steps (see box 2). Typically the learning process is organized in three meeting cycles (Woods 1995). In the first meeting with a new scenario the students work through steps 1 to 5. The second two meetings are devoted to getting feedback on what the students have learnt from the research that they have undertaken between the meetings, synthesizing and applying this information to the scenario. At the end of each cycle the group reviews its performance and learning goals are identified for improvement.

The Teachers Role in PBL

The teacher's role is one 'facilitator of learning' for one or more groups. Facilitation in this context can be defined as playing the role of the more knowledgeable member of the social community of which the student is also a member. Assistance for learning is provided through interactions characterized by such activities as directing, modelling, questioning, and providing cognitive structuring and feedback until the learners are able to perform without assistance (Rideout & Carpio 2001).

Box 1: Example of PBL scenario from Advanced Diploma in Medical Nursing at Middlesex University

Fred Smith is a 62-year-old retired building contractor is admitted to your ward via the Accident & Emergency Department. A CAT scan confirms the diagnosis of a stroke. Three days after admission he has a dense left hemiplegia and remains drowsy and agitated. Mr. Smith has a Grade 2 pressure ulcer on his sacral area. The student nurse reports this to you. You note that there is no record of this in the nursing notes. It has not proved possible so far to insert a nasogastric tube. Mr. Smith's family are distressed about his condition and on the late shift tell you they are worried because Fred is not being fed which will upset his Diabetes.
Scenario also includes a sample of nursing care goals formulated according to Peplau's model and an assessment report from the speech and language therapist

Possible learning areas covered by scenario
Biological: *Neurological observations: use of Glasgow Coma Scales & problems associated with it. Physiology of stroke, Concept of dysphagia: measurement & management
- airway protection, maintaining nutritional status, *physiology of wound healing
properties of wound dressings, principles of stroke rehabilitation: e.g. positioning, preventing hazards of bed rest etc., Nutrition / dehydration
Psychology: Coping with loss, Frustration, Body image
Sociology: Role change, Meaning of illness versus disease
Aesthetic:* Principles of rehabilitation, *prevention and care of pressure sores, *Mouthcare:
*Care of Percutaneous Endoscopic Gastrostomy,
Empirical: Theories of rehabilitation, Information giving,
Professional: *Working within a MDT, Caring for the family,
Ethical: Informed consent

Evidence about the effectiveness of PBL

Norman and Schmidt (1992) argue that there is good empirical evidence to support at least two of the key aspects of PBL in the cognitive psychology literature. Firstly that learning is improved where there is activation of prior knowledge and secondly that elaboration of knowledge at the time of learning enhances retrieval. However with regard to some of the other key aspects of PBL notably self-regulation and group participation Evenson and Hmelo (2000) argue that the theory is a bit vague and that there is a lack of empirical evidence. In addition to this Woodward (1997) highlights the lack of evidence to support the claim that PBL produces practitioners with consistently high levels of performance that are maintained throughout their professional career.

It is claimed that PBL delivers additional benefits in terms of knowledge, understanding, critical thinking, communication, problem solving, teamwork and student satisfaction. Reviews of PBL are difficult to interpret due to the varying methodological approaches used by the reviewers. Three reviews, published in 1993 came to different conclusions. Vernon and Blake (1993) concluded that "results generally support the superiority of the PBL approach over more traditional academic methods". Albanese and Mitchell (1993) concluded that PBL was more nurturing and enjoyable and that PBL graduates performed as well and sometimes better on clinical examinations and faculty evaluations. However they also stated that PBL graduates showed potentially important gaps in their cognitive knowledge base, did not demonstrate expert reasoning patterns, and that PBL was very costly. Berkson (1993) was unequivocal in her conclusion that "the graduate of PBL is not distinguishable from his or her traditional counterpart". She further argued that the experience of PBL can be stressful for the student and faculty and implementation may be unrealistically costly.

Project on the Effectiveness of Problem Based Learning

PEPBL is a three-year research and development project funded by the ESRC Teaching & Learning Research Programme. The project contains two distinct but related empirical research studies. A systematic review of the effectiveness of PBL featuring secondary data analysis of previous studies of PBL that meet explicit pre-specified criteria both for research design and quality and for curriculum design. Secondly a randomized field trial the object of which is to compare the attainment of students in a continuing nursing education programme organized as a PBL curriculum with students in the same programme organized as a 'traditional' curriculum.

Methodological approach

The study can be located under the broad heading of evaluation research. The broader aim of evaluative studies of PBL will be to find out what kinds of PBL produce what learning outcomes for which students in which contexts and to ascertain the relative advantages offered by adopting the PBL approach compared with any other. This study aimed to make a contribution to this agenda by testing the null hypothesis that the use of Problem Based Learning {PBL} curriculum makes no difference to the attainment of nurses undertaking a continuing nursing education programme. Underpinning this approach is the most common form of causal explanation based on four principles (Blaikie 2000):

- There is a temporal order in which cause must precede effect
- There is association that requires that the two events occur together
- There is elimination of alternatives in order to be able to claim that the effect was due to the specified intervention and not something else.
- Causal relationships are made sense of in terms of broader theoretical ideas or assumptions.

In the context of this study the broader social scientific concept of causal mechanism as a set of conditions that when taken together produce an effect informs interpretation of the data (Selltiz et al. 1976). The section below that reports the design and methods used in the study demonstrates how the first three of these principles were met. The search for the broader meaning of these answers will include linking the data to that from other studies of PBL. The interpretation will explore the results in terms of broader ideas about pedagogy and learning.

Research design

The first three of these principles are primarily issues of internal validity and as such are 'managed' through the selection of the research design and the management of the research process. All possible threats to internal and external validity cannot be controlled in any one study, complex educational programs are implemented differently in various settings and are influenced by a host of political and social contexts. For these reasons smaller studies aimed at minimising bias {internal validity concerns) and random error {statistical validity concerns) are valuable in new or innovative educational programmes (such as PBL) (Besson et al. 1982).

A randomised experimental research design was used. Evaluations of study designs have demonstrated that the well designed and executed randomised experiment is superior to any other design at minimising bias and random error and thus is considered most useful to demonstrate programme impact (Boruch & Wortman 1979). The experiment is a particularly efficacious design for causal inference. Random assignment creates treatment groups that are initially comparable (in a probabilistic sense) on all subject attributes. It can then be reasonably be concluded that any final outcome differences are due to treatment effects alone, assuming that other possible threats to validity have been controlled (Tate 1982). The pragmatic trial design used meant that the environment in which the experiment was conducted was kept as close as possible to normal educational practice. There is no placebo or sham intervention and all students who took the programme were included in the evaluation on & (Torgerson & Torgerson 2001).

The disadvantage of the pragmatic trial approach is that there is greater variation making it harder to detect small effects. A number of modifications of the simple two group experimental design were considered to help offset this including 'matching subjects' (Robson 1993), 'repeated measure' or 'cross over' designs (Louis et al. 1984), 'Single subject (A/B)' designs (Robson 1993) and the 'two group pre and post – test' design (Robson 1993). However the way that recruitment to the programmes were organised meant that it was not possible to obtain any data about the participants prior to them starting the programmes. It was also felt unacceptable to ask students to complete any kind of assessment at the beginning of the programme. Given the part-time nature and short duration of the programme it was felt unlikely that the requirements for adequate duration of intervention and washout period required for crossover or single subject designs could be met (Senn 1993).

Evaluating a complex intervention

As the design of the study progressed it became apparent that evidential claims about PBL lacked both methodological and conceptual clarity (Colliver 2000;Maudsley 1999). Furthermore PBL can be considered to be a complex intervention and thus subject to the specific difficulties in defining, developing, documenting and reproducing all such interventions. The Medical Research Council (MRC) framework for the design and evaluation of complex interventions to improve health, is equally applicable to complex interventions in other fields such as education (Campbell et al. 2000). The framework utilises a sequential phased approach to the development of randomised trials of complex interventions. Using this framework the PEPBL study can be considered a phase II exploratory trial. A phase II exploratory trial is concerned with defining the control intervention, estimating the size of the effect, identifying and piloting various outcomes and outcome measures.

Whilst the distinction between exploratory and definitive trials provides a useful framework for study design in practice the boundaries between an exploratory (phase II) trial and a definitive (phase III) trial are blurred. In this study effect sizes and outcomes were identified prior to the study and thus are amenable to hypothesis testing. However given the notable difficulties in measuring the impact of education (Van Der Vleuten 1996) and the lack of valid reliable instruments in PBL, few of the instruments used in the study have been used in studies of the effectiveness of PBL before. This practical blurring of the boundaries also highlights the conceptual blur between the two phases. Given the variety of educational contexts it is questionable whether there could be '*a*' definitive trial of PBL. It maybe that there will need to be definitive trials of PBL in different education contexts of which Continuing Professional Education (CPE) is one

Selection of outcome measures and instrumentation

Cervero's (1988) framework for the evaluation of continuing education for professionals (see box 2). was used to guide the selection of appropriate outcome measures and instrumentation for the study. 'Programme design and implementation' is discussed in detail in part II of the report. The category 'Impact of the application of learning' refers to the so-called second order effects. In the context of this study this refers to whether there are measurable improvements in patient outcomes as a result of nurses undertaking a continuing nursing education programme. Measurement of such effects was beyond the scope of this study.

Box 2: *Framework for the evaluation of continuing professional education (Cervero 1988)*

- Programme design and implementation
- Learner participation
- Learner satisfaction
- Learner knowledge, skills and attitudes
- Application of learning after the programme
- Impact of application of learning (second order effects – e.g. improvements in the health of patients

Given the methodological approach of the study and the limited time and resources available effort was made to identify existing sensitive, valid and reliable outcomes and instrumentation that would achieve high response

BEST COPY AVAILABLE

rates. The setting of the experiment i.e. as a pragmatic trial in a 'real world' education setting provided an additional set of constraints. Any research measurement needed to place as little burden on the students and teachers as possible and not to divert students from learning. It was therefore agreed that it would be unreasonable and impractical to require students to undertake any additional form of summative testing or assessment. The outcomes selected and instrumentation used are summarised in tables 1 & 2

Table 1: Summary of outcome measures and instrumentation (excluding learner outcomes)

Category	Measure
Programme design and implementation	Tutor record of session content and activity Interaction analysis Non participant observation Participant observation
Learner participation	Tutor records of student attendance activity Interaction analysis Student study workload (self reported)
Learner/teacher satisfaction	Course Evaluation Questionnaire Observation Teachers Diaries Nominal Group technique Drop-out rates Exit Interviews Follow-up questionnaire
Application of learning after the programme	Follow-up questionnaire of students Follow-up questionnaire of students' managers

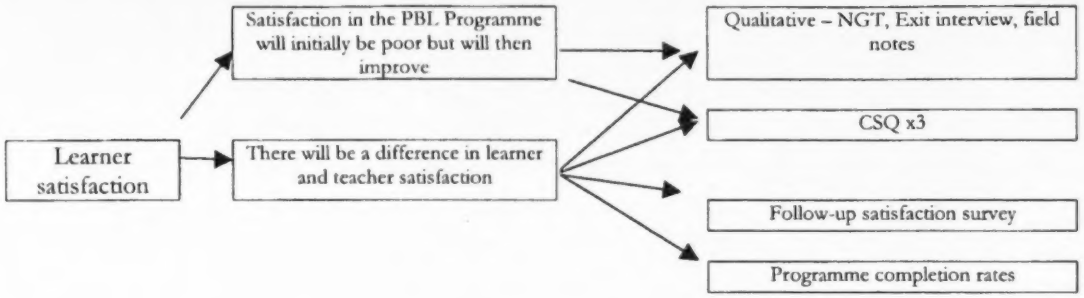
Framework category Learner Participation

Differences in learner participation in the two curricula and within the classroom are discussed in detail in part II of the report. Another focus was students workload which can be a useful as curriculum evaluation tool (Swanson et al. 1991), providing a proxy indicator of programme quality (Snellen-Balendeng & Schmidt 1990). The Student Workload Questionnaire developed specifically for the study required students to report the amount of programme related work undertaken in the week prior to the administration of the questionnaire. This approach has obvious limitations in that students are being asked to recall activity and may be prone to over or under reporting. Additionally the timing of workload requirements is likely to vary between teachers and between curricula. For these reasons the questionnaire was completed five times by each student at randomly selected points during the academic year. Analysis compared the average and variation in self-reported workload in each curriculum.

Framework Category Learner satisfaction

It is often claimed that PBL leads to increased levels of learner satisfaction or that students like PBL (Wilkie 2000). This would seem to be an important outcome both for its own sake and because of an imputed link between enjoyment, motivation and performance (Mckeachie et al 1986). There are a number of ways of conceptualising enjoyment and satisfaction in an educational context and therefore a 'basket of indicators' approach was adopted. In this approach the same outcome is 'measured' using a variety of approaches/ instruments (see figure ?). Questions about students satisfaction were also included in the follow-up survey which is discussed in more detail later.

Figure 2: Basket of indicators for the Outcome Learner Satisfaction


Satisfaction in the PBL Programme will initially be poor but will then improve
Qualitative – NGT, Exit interview, field notes
CSQ x3
Learner satisfaction
There will be a difference in learner and teacher satisfaction
Follow-up satisfaction survey
Programme completion rates

The Course Experience Questionnaire (CEQ) (Ramsden 1992) was developed on the basis of empirical and theoretical work on the quality of teaching in higher education. Students are asked to rate the quality of their programme using questions with a five point likert scale. The assessment covers five categories; teaching, goals, workload, assessment and student independence . The CEQ was tested in 50 Australian education institutions on 4500 students cross a range of disciplines and was found to discriminate between teaching styles and quality within and between different subject areas (Ramsden 1992). The use of the CEQ is now compulsory in Australian Higher Education Institutions (Long & Johnson 1997). The CEQ was also used to evaluate student satisfaction on the Problem Based Learning Programmes in the Health Science Faculty at Griffith University in Brisbane (Margetson 1995). The CEQ has been updated several times. One reason for using the original version of the CEQ is that The scale 'Emphasis on independence' has been dropped from more recent versions of the scale now in widespread use (Long & Johnson 1997). It was felt that this scale might be highly appropriate to for identifying differences between student's perceptions of PBL and non- PBL courses.

There are a variety of Nominal Group Techniques (Delbecq & Van den Ven 1971). The approach used in the study was a variation of the RAND form of NGT (Black et al. 1998). The instructions given to students are shown in box 3. The NGT was undertaken on the final day of each groups programme. The instructions were given to the students by the Principal Investigator. The Principal Investigator and the teacher left the classroom until the students had completed the exercise. After the students had completed the exercise the lists generated by the students were discussed with them to gain greater clarification.

Box3 Instructions for Nominal Group Technique

1. List five things that you have enjoyed about the programme

2. List 5 things that you found difficult on or about the programme

3 After all the group has completed parts 1 & 2 compile a group list using the items highlighted by each individual eliminating any duplications

4. Each member of the group has 5 points to award to the things that they enjoyed most from the group list. You can allocate the points in any way that you choose. For example you could allocate all points to one item or 3 points to one and two to another or 1 point to each of five different items. You do not have to give the points to the items that you chose originally, if you feel that there are other items on the group list that are more important.

5. Each member of the group has 5 points to award to the things that they enjoyed least from the group list (5 = least enjoyable). You can allocate the points in any way that you choose. For example you could allocate all points to one item or 3 points to one and two to another or 1 point to each of five different items. You do not have to give the points to the items that you chose originally, if you feel that there are other items on the group list that are more important.

6. Add up the points on the list to arrive at 5 best and 5 worst things on the course. From the perspective of the group

Telephone exit interviews were carried out with all students who discontinued the programme for whatever reason. The interview schedule was designed specifically for this study. Students were contacted as soon as the

BEST COPY AVAILABLE

Principal Investigator became aware that they had left the programme. The period of time between the students last teaching session and when they were contacted varied as it was often not confirmed for some weeks that a student had actually quit the programme as opposed to just being absent. The Principal Investigator contacted the student to arrange a convenient time for the telephone interview. During the interview the Principal Investigator made note of the students responses and wrote up the interview immediately after the interview was complete. Analysis of the exit interviews was carried out by the Principal Investigator and comprised of reviewing the completed exit interview schedules to identify areas of commonality and difference in the students accounts.

Framework category changes in learner knowledge, skills and attitudes

This category focuses on measuring changes in the learner's cognitive, affective or psychomotor competence (Cervero 1988). Despite the extensive literature on assessment of professional competence there is little consensus about what exactly should be measured let alone how it should be measured. (Van Der Vleuten 1996). An important aspect of PBL philosophy is the recognition of the fact that assessment has a major impact on learning. However, there is not a consensus on the either the outcomes or methods of measurement that should to be used to evaluate the effects of PBL on student knowledge, skills and attitudes. A range of student capacities under this heading can be identified in the PBL literature. A summary of the claims made for PBL produced by Engel (1991) was used to guide the selection of appropriate outcome measures and instruments in this category. The claims/ goals of PBL and the approach taken to its 'measurement are summarised in table two The selection and use of measurement tools for the study involved a trade off between reliability, validity, educational impact, acceptability and cost which are discussed detail below.

Table 2: PBL claims and their respective study measurement instruments

Claim / PBL goals	Outcome / instrument
Adapting to and participating in change	Assignment Follow-up questionnaires
Dealing with problems and making reasoned decisions in unfamiliar situations	Assignment Group work video assessment Follow-up questionnaire
Reasoning critically and creatively	Assignment Group work video assessment Follow-up questionnaires
Adopting a more universal or holistic approach	Assignments Follow-up questionnaires
Practising empathy/ appreciating other persons point of view	Assignments Group work video assessment Follow-up questionnaires
Collaborating productively in groups or teams	Group work video assessment Follow-up questionnaires
Identifying own strengths and weaknesses and taking appropriate remedial action	ASSIST

Reliability of assessment instruments

The key problem identified in research on performance assessment is the variability of candidate performance on even very similar cognitive tasks. This occurs whatever the competence being measured and whatever response format is used (with the possible exception of Multiple Choice questions containing a large sample of items) suggesting that assessments containing a small sample of items e.g. essays, produce unstable or unreliable scores (Swanson et al 1991). Van Der Vleuten(1996) argues that the practical consequences of this are that the sample size of test items should be sufficiently large and the test designed such that the affect of variability on the precision of the instrument is minimised.

Validity of assessment instruments

The assessment of validity i.e. that tests measure what they are required to measure, requires the identification of good criteria or standards. In most areas of professional competence good criteria and perfect standards do not exist (Van Der Vleuten 1996). PBL is no exception. There is no agree approach for example to measuring critical thinking skills. A recent evaluation of the Problem Based BSc Nursing programme at McMaster University in Canada included use of 'The California Critical Thinking Skills Test' (CCTST) (Facione 1990), (Personal communication Liz Rideout). The CCTST is based on the consensus view of the critical thinker produced by the American Philosophical Association and has undergone extensive testing by the authors (Howell Adams et al. 1996). Numerous criticisms have been made of both the CCTST but they are probably as useful as any other standardised critical thinking test Howell et al 1996}. However, the main problem of all such tests lies in the way that critical thinking is conceptualised independently of context. Fisher and Scriven (1997) argue that critical thinking is underpinned by informal logic, and is thus context dependent. PBL is based on principles derived from cognitive psychology i.e. that knowledge is structured in semantic networks. PBL scenarios create a semantic structure for the learning of knowledge which is similar to the semantic structure in which the knowledge will be applied thus enabling the recall of required knowledge (Gijselaers 1996). It would therefore seem 'invalid' to use context free critical thinking tests to measure outcomes achieved by PBL.

Another 'validity' issue in relation to PBL is the shared view amongst PBL advocates that assessment drives learning. However the consequences of this view are interpreted differently. Some writers suggest that both the response format and the content of the test must be appropriate to PBL (Marks-Maran & Gail Thomas 2000). Others argue that response format is of less consequence than content and test-design (Norman 1991). The Multiple-Choice Question format (MCQ) was introduced to cope with the increased logistical demand for educational testing and to provide reliable assessment of student performance. MCQs have often been rejected for use in PBL programmes for various reasons including the belief that they are only suitable to measure lower levels of taxonomic cognitive functioning (Van Der Vleuten 1996). However others argue that there is no reason why MCQ cannot be used in PBL assessment as the key issue is the quality of the design and administration of the test rather than the method itself (Swanson et al 1991). Moreover the MCQ is used for assessment of student performance with slight variations on the Medical PBL programmes at all American Medical Schools (a licensing requirement), the PBL programmes in Medicine at McMaster (Canada) and Maastricht (Netherlands) Universities.

There are a number of assessment formats that are claimed to provide more valid assessments of the learning developed by PBL programmes. Modified Essay Questions (MEQ) have been used to stimulate problem-based learning in both clinical and pre-clinical courses. It is argued that the properly designed evolving MEQ opens up possibilities for exercising 'intelligent guessing' that mirrors the realities of clinical work and can thus measure abilities and attitudes that other assessment methods cannot (Knox 1980). Although the reliability of the MEQ method has been established (Feletti 1980) caution has been expressed about its misuse and over use in PBL programmes (Felletti & Smith 1986). Studies have also suggested that the MEQ measures nothing different from the MCQ (Norman GR et al. 1987). MEQs are used as part of the assessment programme on the BSc Nursing Programmes at Thames Valley University and the University of Dundee which both use Problem Based Learning. (Marks-Maran & Gail Thomas 2000). However the reliability of these MEQ's has not been established. This and practical constraints prevented their use in this study.

The Triple Jump Exercise (TJE) is a learning process measure widely used as an assessment tool in PBL programmes (Painvin et al. 1979). The TJE consists of three steps (jumps.) A structured oral examination based on one or more patient problems, a time limited study assignment in relation to the patient problems in the first oral and a repeat oral examination in which the quality of self – learning around the assigned topic is assessed. The TJE is currently used in a number of PBL programmes around the world, including the Problem-based BSc Nursing programme at McMaster University in Canada. The TJE is a very time consuming, costly method of

assessment with poor measurement characteristics (Blake et al. 1995). These factors combined with practical constraints prevented the use of the TJE in this study.

Existing programme assignments – Free response format

The written assessment methods currently used in the programme use the free response format (see box 4). With their emphasis on self selection of topic, self-directed information searching and presentation of data in a clear focussed manner, written assignments are viewed as a relevant evaluation method within the PBL approach (Rideout 2001). They are widely used in assessment programmes on PBL courses (Marks-Maran & Gail Thomas 2000). The pre-existing course assignments are congruent with the aims of PBL and have the advantage that the students would be motivated to complete the assignments well given that they are a programme requirement. It was therefore decided that students' assignment scores should be used as one of the outcome measures for the research study.

Box 4: Written assessments used on advanced diploma programme

- Literature review and seminar presentation;
- Care study and supporting essay;
- Learning contract and reflective account

However, the poor intra and inter observer reliability of marker evaluations of free response assessments are well documented (Biggs 1999;Brown et al. 1997;Swanson D 1987;Van Der Vleuten 1996). Analysis of available data on assignment scores from previous years of the participating programme reveal a skewed distribution towards the higher end of the marking scale which did not match the teachers verbal accounts of the performance of previous students. It can be argued that the cause of these validity and reliability problems is the tutors marking rather than anything inherent in the method itself (Swanson et al 1991). The provision of simple protocols to structure and score oral examinations can significantly improve the reliability as compared to free judgement (Verma M & Singh 1997).

Minimising observer bias – External, independent blind marking

There is evidence that unblinded outcome assessment, particularly for subjective outcomes (such as used here), is demonstrably associated with bias (Prescott et al. 1998). The assignment scores used for the research were therefore generated independently from the marks given by teachers to meet the programme assessment requirements. Three nurse teachers from other UK universities with experience were recruited to mark the assignments. Each marker was a nurse and had experience of teaching and marking in pre and post registration programmes. The markers had no previous connection with either Middlesex University or any member of the teaching or research team in the study. The markers were paid the standard University external examiner fee. The scripts were anonymized by removal of all identification except a student number, and sent to the external examiner by post for marking. The marking for research purposes was therefore carried out by independent experts, 'blind' to the allocation status of the students.

Improving the reliability and validity of the expert marking

Despite agreement that marking protocols are useful there are huge variations in the types of protocol used and disagreement about the nature of the criteria that should be included. According to Biggs (1999) this is partly due to different views about 'learning' and assessment and also because of the dominance in higher education of the norm referenced approach to assessment. He argues that this often results in marking protocols that do not reflect what it is the 'teaching' is trying to achieve, either through omission or through the use of an analytic approach in which the big picture of performance is somehow lost in the detailed criteria. Detailed criteria have

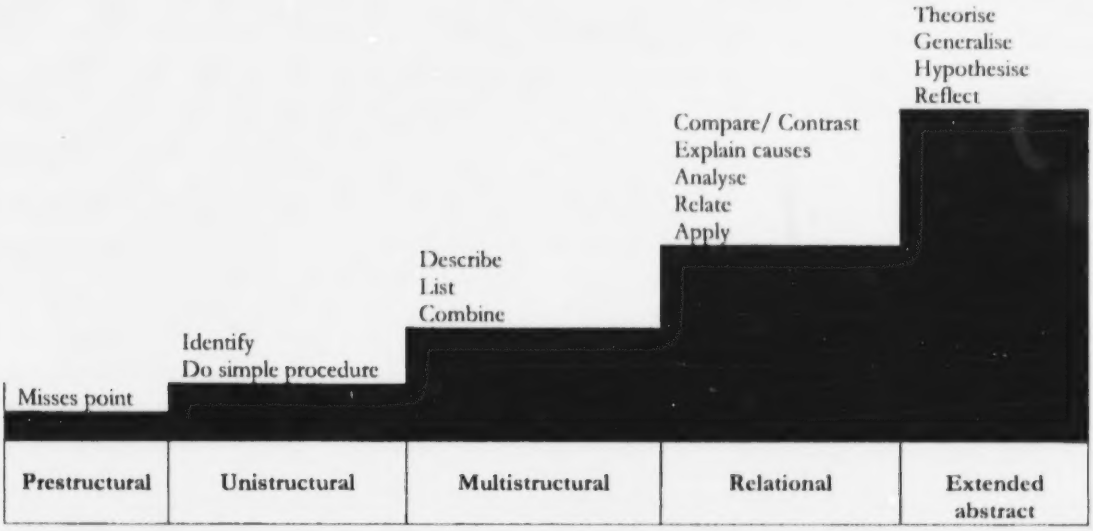
BEST COPY AVAILABLE

been showed to yield more to low level learning i.e. students can obtain high marks even though only lower level learning has been demonstrated and also fail to improve reliability due to their difficulty in use (Brown et al 1997). However more detailed criteria can be useful for research purposes, but only to the extent that markers will actually use them.

The purpose of the programme assignments is to measure the extent to which a student has achieved the objectives or learning outcomes of the programme. The aim for the new protocols was firstly to ensure that what is marked reflects the programme objectives i.e. is valid. With respect to this point it should be noted that it was not the intention to develop new or different criteria that did not reflect the course objectives or the information that students were given. This would be of questionable validity. Secondly to improve reliability i.e. the likelihood that the same person would make the same judgement about the same performance on two different occasions (intra-observer reliability) and different judges would make the same judgement about the same performance on the same occasion (inter-observer reliability).

In relation to validity the issue is to ensure that understanding is defined in ways that do justice to the topic/ content taught and level of study as exemplified in the in programme objectives (Biggs 1999). The SOLO taxonomy provides a general framework for structuring levels of understanding . It is based on the study of student outcomes in a variety of academic content areas which demonstrated that as students grow the outcomes of their learning display similar stages of increased structural complexity (Biggs 1999). Levels of understanding can be described as verbs in ascending order of cognitive complexity that parallel the SOLO taxonomy (see figure three) (Biggs 1999).

Figure 3: The SOLO taxonomy and hierarchy of verbs that indicate increasing cognitive complexity



An analysis of the programme objectives and assignment information given to students identified that most of the verbs used are firmly in the relational stage of the taxonomy extending in some parts to the extended abstract level. The purpose of the assessments as stated in the student handbook is given as "to reveal the student's ability to synthesise and evaluate the theoretical issues of each of the modules and to facilitate student's exploration of their value system which underpins their professional practice" . The requirement for this level of understanding is congruent with the final year undergraduate, status of the programme. The marking protocol improves reliability by identifying clearly and unambiguously what the marker should be looking for in terms of level of understanding displayed in the students writing and how these components should be weighted when considering the overall mark allocated.

The new marking protocols were based on existing standard models (Brown et al 1997;Johnson 1993) The qualitative description of each category was modified to reflect the SOLO taxonomy and the specific

requirements of the assignments in particular the relation of theory to practice.. Guidelines on the process of marking were also been provided (see box five) to minimise halo and systematic order effects (Biggs 1999;Brown et al 1997).

Box 5: Marking process recommendations for external markers (Biggs 1999;Brown et al 1997)

- Mark intensively until you have the criteria fixed in your head, then you can mark reliably a few questions at a time between other tasks
- At the beginning of each marking session (if there has been a gap since the previous session) re mark a few scripts
- Grade coarsely at first (qualitatively) by skim reading all the scripts and place in piles according to criterion categories. Then re-read with the criteria and mind to give quantitative value. Be prepared to change scripts at the borderlines of each category
- Shuffle the scripts between first and second readings
- Use the whole range of grades between 0 and 100%

Measurement of capability to practice empathy and collaborate productively in groups.

The goal of practising empathy was considered as part of the goal of collaborating productively in groups. PBL places great emphasis on group or teamwork. It is argued that the process of collaboration improves the effectiveness of learning and the effectiveness of the individual in future collaborative settings (Myers Kelson & Distlehorst 2000). The claim that PBL improves group work skills and that this improvement produces measurable increases in leaning and thinking and later on in patient care appears to be an assumption that requires further testing by research (Thomas 1997). Given the importance attached to groupwork, there appears to be a deficit of rigorous evaluative studies of group performance in the PBL literature.

An attempt was therefore made to assess this aspect of student performance using video assessment of each group undertaking a series of problem solving tasks. The studio facility used was based on one of the University sites. The video assessment was carried out on the last day of each groups programme. The groups were informed in advance that the exercise was being conducted. On the day each group was taken into the studio facility. The group sat in a semi-circle with small desks for each group member and a flip chart and pens were made available. The audio-visual technicians provided a briefing on the technical aspects of the recording process and visual and sound checks were undertaken. The Principal Investigator gave a briefing and instructions to each group. Identical instructions were given on each occasion. The Principal Investigator watched the groups from the studio control room and interrupted groups only if they violated any of the rules laid down for each problem solving exercise. The video was recorded onto a master tape using one fixed and two roving cameras. The Principal Investigator and control technician selected shots from the live feed. The master tape was then edited onto a VHS tape showing each group performance in full.

The problem solving exercises were compiled from problem solving texts. The exercises were selected to provide a mixture of paper based and physical problems that were not directly related to the participants workplace. The problems also vary in the extent to which they require logical, practical and/or spacial awareness. It should be noted that the exercises were not designed specifically to test problem solving ability but rather to stimulate the group to use its collective skills/ knowledge/ abilities to solve the problems i.e. to perform as a group. The exercises were not formulated as clinical 'scenarios' or triggers or problem solving exercises in order to minimise any advantage that the PBL groups might have due to their previous experience of these kinds of exercises or 'cueing'.

The task of evaluating how well a team or group functions could be viewed as a simple task of measuring how effective a group is at achieving the objectives that it is set. However the real world is rarely as simple as this as groups are dynamic, tasks vary in complexity and groups work in different and complex contexts. The literature on group work assessment has therefore focused on identifying the kinds of activities/ characteristics/behaviours/ attitudes which individuals in groups and groups themselves need to develop to

BEST COPY AVAILABLE

perform successfully in the complex settings. Developments in measurement have proceeded alongside the identification of these characteristics.

Attempts were made to identify tools that can be used to measure how effective a group is at working together both in the PBL literature and more widely in the literature on group work. Within PBL the majority of instruments identified are primarily for the use of group members themselves in the process of evaluating group performance for formative purposes e.g. Group enrichment task (Woods 1995) or the small group teaching evaluation used at McMaster University (Jaques 1990). In the broader literature on group work, other instruments were identified that help individuals/groups identify the roles that members take in groups e.g. Team Orientation and Behaviour Inventory (Goodstein et al. 1983) and/or how they view each others behaviour e.g. The interpersonal perception scale (IPS) (Patton et al. 1989).

The Faculty of Medicine and Health Sciences, Newcastle University, Australia developed an observational assessment tool that is used both formatively and summatively to assess group process and group reasoning (Rolfe & Murphy L 1994). The instrument is used to observe group performance during a specific group task and is carried out in two stages. The instrument consists of 22 criteria in 3 domains. No data is reported on reliability and or validity and contact with the authors confirms that no subsequent evaluation of the instrument has been carried out (I Rolfe personal communication November 2001}. Each criterion is specified as a pair. The first behaviour is that which is considered appropriate, the second that which is considered inappropriate. The instrument also offers the possibility of assessing other outcomes of PBL namely 'Dealing with problems and making reasoned decisions in unfamiliar situations and 'Reasoning critically and creatively' (Engel 1991). The nature of the assessment task set for the groups in this study meant that it would not have been possible to make judgements about all the criteria on the original instrument. So only the relevant items were included in the version used.

Two independent 'experts' carried out the assessment of the video footage using the instrument. One was a social scientist with experience of group observation techniques. The other was a professional training consultant whose training activity included providing training on team/group work. Neither had any experience of PBL. The assessors were provided with an edited VHS video to analyze 'at home' independently of each other. Groups were identified on the video with a number. The assessors were therefore 'blind' to the allocation status (i.e. experiment or control) of each group.

With hindsight it seemed likely that problem solving exercises in multiple solutions and which may involve the making of value judgements were more likely to provoke behaviour that revealed a groups capabilities at as working together. It was also unrealistic to require assessors to analyse more than 10 hours of video footage. It was therefore decided to focus the analysis only on the problem solving tasks that appeared to provoke the most discussion / non- consensual debate amongst the groups. The Principal Investigator reviewed all the video footage and three problems were identified in this category, 'The bomb scare', 'The line problem', and 'Build a bridge'. In the year two videos because the groups had been set a time limit for completion of all the exercises and these three problems were completed in approximately 15 minutes. They were therefore included on the assessors edited video in their entirety. No time limit was given to the first year groups and therefore they took longer to complete the exercises. In order to bring the length of video footage for these groups down to roughly the equivalent of the year 2 groups the video footage of these problems was edited to remove excess periods of silence or inactivity.

Assessing PBL Goal: Improving self – directed learning

One of the most influential concepts in higher education is that of 'learning styles' (Kolb 1984) or 'approaches to learning' (Ramsden 1992). (The term approaches to learning will be used here). It is argued that learning comprises both of what we learn and how we learn it. There two ways in which learning can take place;

Holistically or atomistically. What we learn can be assessed in terms of the meaning or significance of the task. 'Deep' learning focuses on what the task is about and 'surface' learning focuses on the signs e.g. remembering dates. It is argued that research has demonstrated that 'Holistic' 'deep' learning is more successful than 'atomistic' 'surface' learning for understanding (Marton et al. 1984). It is argued that students who adopt these less effective learning styles can be identified and remedial action taken (Tait & Entwistle 1996). Similarly observation of how students study can result in a useful indicator of the learning processes that occur (Coles 1990).

ASSIST (Approaches and Study Skill inventory for Students) (Tait & Entwistle 1996) was developed from the approaches to studying inventory ASI (Entwistle & Ramsden 1983). Both ASSIST and the ASI have been used in large numbers of studies including studies of PBL (Coles 1985). Three or four factors typically emerge from item analysis which represent deep, surface, strategic (equivalent to holistic above) and apathetic (equivalent to atomistic) approaches to studying. Relationships with academic performance are also fairly consistent with positive correlation normally found with the strategic approach and negative correlation's with both surface and apathetic approaches (Tait & Entwistle 1996). The short version of ASSIST that focuses on approaches to studying and preference for different types of course or teaching was used. The instrument was administered to participants at the beginning of the programme and again on completion of the programmes. Analysis will focus on comparing the difference in the changes between the groups.

Assessing application of learning after the programme

Consideration of the long term effects of any educational programme is an important aspect of measuring programme impact (Wilkes & Bligh 1999). The question is whether improvement on some kind of assessment immediately on completion of the educational intervention actually translates subsequently into improved performance (Abrahamson 1984). The issue is particularly pertinent where the educational programme has a direct vocational role i.e. the preparation and/or continuing development of practitioners in a particular field. It is quite possible that the impact of learning on practice may not become apparent to the learner (or the external observer) until some period after the conclusion of the educational programme (Pascarella & Terenzeni 1991). Consequently follow-up studies may produce quite different results to those obtained at the immediate completion of the programme. Claims for the importance and /or legitimacy of Problem Based Learning (PBL) usually emphasize the need to develop new kinds of practitioner, improve the performance of practitioners and/or improve student satisfaction (Albanese & Mitchell 1993). The technical and methodological difficulties of assessing impact at this level of complexity coupled with the limited duration and funding of most educational evaluations means that there are comparatively few studies of this kind (Hutchinson 1999). The limited resources available to the project meant that the only possible method of data collection for the longer term follow-up was a postal survey. It is argued that six months is a period of time in which the quality of opinion about the utility of the programme is more likely to be experience based and less likely to be based on factors such as entertainment or prestige (Nowlen 1988).

No follow-up studies of PBL that were valid, reliable and relevant to this students group and context were identified that could be used for this study therefore 'new' instruments were developed specifically for use in this study. The measurement instruments used were embedded in a questionnaire designed for use in a postal survey. Consideration was given to ease of and time for completion in order to minimise the likelihood of non-response and the return of incomplete questionnaires. A structured format that in the main uses predetermined standardised response formats was selected to aid completion, increase reliability and facilitate data analysis. The questionnaire for former students' comprises; questions about any changes in their work role since completion of the programme, a set of statements designed to assess their performance, a set of statements designed to obtain their views about the impact of the programme on their practice and a set of statements designed to assess their views about the strengths and weaknesses of the programme.

It was recognised that performance in these areas is interlinked both conceptually and in practice and furthermore that assessing performance in areas such as these areas is highly problematic (Hutchinson 1999). A multi-item scale was created to assess performance in each dimension. Each scale used a number of items that were developed from tools used in previous studies of PBL impact (Peters et al. 2000;Walton J et al. 1997;Woodward & Ferrier 1982) and from other relevant performance assessment tools (Brown et al 1997;Patton et al 1989;Quinn et al. 1990;Redding 1992).

Pilot line managers questionnaire

As a form of triangulation student's immediate line managers were asked to rate their performance. The students varied with regard to their position in the organisational hierarchy, for example, some were ward managers and others junior staff nurses. This suggests that the person who has 'line managerial' responsibility for a particular participant will not always work with them sufficiently closely to be able to provide an assessment at the same level of detail as that required by the instruments in the student questionnaire. The multi-item assessment instrument used in 'the line manager' questionnaire was developed from other tools used to assess performance of students in work related behaviours (Brown et al 1997;Patton et al 1989) that the educational programmes in this study claim to develop.

Pre-testing of pilot student follow-up questionnaires

Pre-testing of the questionnaire broadly followed the procedures outlined by the American Statistical Association (1997). The paper outlining the development of the questionnaire and the questionnaire itself were made available from the project website and the project e-mail list used to ask for comments and feedback. The questionnaire was redrafted as the result of a small pilot study and the identification of further relevant literature. The questionnaires underwent several revisions as a result of the identification of new literature and 2 rounds of pretesting with students and managers not involved in the programmes being investigated in this study.

Development of 2nd version of questionnaires

The combination of internal and external review, the identification of other relevant literature and results of pre-testing indicated that substantial modification to the student questionnaire was required. A systematic review of research evidence and best practice in questionnaire design became available in early 2002 (McColl & Jacoby A 2001). On the sub-scales teamwork, leadership and clinical practice the removal of items with low Alphas scores and/or with possible confusing negative wording left 21 items remaining. These were revised into a single 21 item scale measuring the dimension 'Capability for Clinical Practice Organisation'.

The Self Directed Learning Readiness Scale (SDLRS) (Fisher et al. 2001) measures the degree to which an individual possess the attitudes, abilities and personality characteristics necessary for self directed learning. The instrument was developed by nurse educators in Australia using a rigorous three stage process. In the first stage a bank of 93 items were developed from the existing literature. In the second stage a 2 round modified Delphi technique was used in which selected experts independently identified those items that they felt were necessary for self-directed learning. In the third stage pretesting of the SDLRS the final selection of items was carried out using item-total correlation based on data from a sample of 201 nursing students. Items with a corrected item-total correlation score of <0.3 were removed removed from the scale leaving a 40 item scale with an alpha for the total item scale of 0.924. Factor analysis identified three component subscales, Self Management (SM, n= 13 items, $\propto$ 0.857), Desire for Learning (DL, n= 12 items, $\propto$ 0.847), Self Control (SC, n=15 items, $\propto$0.830). Based on the pilot study results the authors argue that a score of 150+ indicates a readiness for Self Directed Learning. The SDLRS instrument was included in the revised student questionnaire.

The questionnaire for managers/supervisors was also revised using the systematic review referred to above. The scale was remodelled to include additional items from the Clinical Supervisors report form developed to assess practice performance of medical students in the PBL programme at the University of Newcastle (NSW) Medical school (Saunders et al. 1982).

Table 3: Student follow-up questionnaire pretesting of final version - Cronbach's ∝ coefficients

Dimension/ Scale	No. Cases	Cronbach's Alpha
'Capability for Clinical Practice Organisation'.	20	0.7518
Self Directed Learning Readiness Scale	22	0.9156
Impact on my practice Scale	21	0.8588
Programme strengths & weaknesses	21	0.8398

The results of the analysis of the internal consistency of the four different scales are given in table three. The consistency of all the scales appears satisfactory. Further analysis of the scale 'clinical performance' revealed that Alphas for the two groups were quite different with the Alpha for one group being 0.5399 and the other 0.8582. For this reason it was decided that the scale would be not be modified further. The total score for the 'Programme strengths and weaknesses' scale showed a statistically significant positive correlation with the students overall rating of how they learnt on the programme (r=0.637 – Significant at 0.01 on a 2 –tailed test), providing some evidence of the validity of the scale items.

Administration of the follow-up questionnaires

The questionnaires were sent to all students who completed the programmes and to the person who they named as their line manager at the time the questionnaire was sent. Each student was contacted prior to the questionnaire being sent to inform them that the questionnaire was being sent and to check that the contact details for them and their line manager were up to date. The questionnaires were sent by post with a personalised covering letter and a prepaid return envelope to maximise response rates. Where possible non respondents were contacted by telephone and additional reminder questionnaires were sent where required. The questionnaires were sent to the first cohort of students approximately 8 months after they completed their programme. The questionnaires were sent to the second group students approximately 4 months after they completed their programme.

Economic evaluation

An important consideration in the evaluation of any teaching and learning strategy in a climate where the resources available for the creation of learning are scarce relative to the demands made upon them is the relative costs of any benefits obtained from using a particular strategy. The basic framework for economic appraisal is that all interventions require resources that have an alternative uses and therefore involve a sacrifice of benefit elsewhere (cost). At the same time all effective interventions achieve results that are of value (benefits). The process of weighing gains against sacrifices is known as the cost-benefit approach (Drummond 1980). Obviously different perspectives can be taken on what is a cost and what is a benefit. To a teacher a finding that students do more 'homework' may be viewed as a benefit, whilst to students themselves this may be viewed as a cost. Obviously the value of any cost benefit analysis is only as good as the data upon which such estimates are based. Data on a range of student outcomes that can be construed as 'benefits' are being collected. Given that one of the major concerns expressed in the PBL literature is that 'PBL is more 'expensive' (Berkson 1993) the 'cost' focus will be a comparison of teacher 'workload' between the two curricula. All the teachers involved in the study contributed to the development of the experimental (PBL) curriculum and the control (SGL) curriculum was already in existence. Therefore the focus of the data collection was on teacher 'workload' associated with the

delivery and support of students during the programme and more specifically during 'term' time. The tutors were provided with a form to record programme associated workload on a weekly basis (see appendix?). Initially teachers were e-mailed on a weekly basis to remind them to complete their forms. However this proved counterproductive as it irritated the teachers. It is likely that the teachers did not complete these forms contemporaneously.

Conclusions

The study was completed in January 2003 and data analysis is currently underway. The study aimed to tackled the issue of research design and assessment of impact in as rigorous a fashion as possible. The use of the randomized experimental design whilst having a long and honourable history in education research (Oakley 2000) is fairly rare in British Higher Education research and its use in research in Problem Based Learning stuides is largely limited to studies of undergraduate medical education. During the course of the study the study design developed and changed as 'new' information became available to the researchers, and as other practical obstacles were negotiated. The end result was a study limited by a series of compromises in its design and execution. However such a fate awaits any real world research whichever design and methodological approach is used. . The analysis will attempt to understand what the limitations of the research are and how the results can be interpreted but we are confident that the study will make a useful and lasting contribution to our understanding of the effectiveness of PBL.

Contact author

Mark Newman. ESRC Research Fellow & Principal Investigator
Project on The Effectiveness of Problem Based Learning (PEPBL)
School of Life Long Learning & Education & School of Health & Social Sciences
Middlesex University
Furnival Building
Archway Campus
2-10 Highgate Hill
London N19 5LW

Tel: 020 8411 6702
e-mail: m.newman@mdx.ac.uk
Project website : http://www.hebes.mdx.ac.uk/teaching/Research/PEPBL/index.htm

References

Abrahamson, S. 'Evaluation of continuing education in health professions: The state of the art', *Evaluation and the health professions*, vol. 7, no. 1, (1984), pp. 3-23.

Albanese, M. A. & Mitchell, S. 'Problem-based Learning: A Review of Literature on its Outcomes and Implementation Issues', *Academic Medicine*, vol. 68, no. 1, (1993), pp. 52-81.

American Statistical, A. *How to conduct pretesting* (Alexandria, ASA, 1997)

Barrows, H. S. 'A Taxonomy of Problem-based Learning Methods', *Medical Education*, vol. 20, (1986), pp. 481-486.

Barrows, H. S. 'Foreword', in *Problem based learning: A research perspective on learning interactions*, D. Evenson & C. Hmelo, eds. (Mahwah, Lawrence Erlbaum, 2000) p. vii-ix

Bereiter, C. & Scardamalia M , 'Process and product in problem based learning', in *Problem based learning: A research perspective on learning interactions*, D. Evenson & C. Hmelo, eds. (Mahwah, Lawrence Erlbaum Associates, 2000) pp. 185-195

Berkson, L. 'Problem-based learning: have the expectations been met?', *Academic Medicine*, vol. 68, no. 10, (1993), p. S79-S88.

Besson, J., Hardin Best, J. H., & Michael, W. B. 'Basic principles of design; analysis in evaluation research,A.twenty-year perspective on evaluation study design I', in *Encyclopaedia of education research*, H. Mitzel & W. Rabinowitz, eds. (London, Collier MacMillan, 1982)

Biggs, J. *Teaching for quality at University* (Buckingham, Society for Research in Higher Education/ Open University Press, 1999)

Black, N., Murphy, M., Lamping, D., McKee, M., Sanderson, C., Askham, J., & Marteau, T. 'Consensus method development for clinical guidelines', in *Health Services Research Methods: A guide to best practice*, N. Black et al., eds. (London, BMJ Books, 1998) pp. 199-211

Blaikie, N. *Designing Social research* (London, Polity Press, 2000)

Blake, J., Norman GR, & Smith, E. K. M. 'Report card from McMaster: Student evaluation at a problem-based medical school', *Lancet; :* vol. 345, (1995), pp. 899-902.

Boruch, R. F. & Wortman, P. M. 'Implications of education evaluation for education policy', in *Review of research in education*, Berliner D.C, ed. (Washington DC, American Educational Research Association, 1979) pp. 309-361

Brown, G., Bull, J., & Pendlebury, M. *Assessing student learning in higher education* (London, Routledge, 1997)

Campbell, M., Fitzpatrick, R., Haines, A., Kinmouth, A., Sandercock, P., Spiegelhalter, D., & Tyrer, P. 'Framework for the design and evaluation of complex interventions to improve health', *British Medical Journal*, vol. 321, (2000), pp. 694-696.

Cervero, R. M. *Effective continuing education for professionals* (San Francisco, Jossey-Bass, 1988)

Coles, C. R. 'Differences between conventional and problem-based curricula in their students approaches to studying', *Medical Education*, vol. 19, (1985), pp. 308-309.

Coles, C. R. 'Evaluating the effects curricula have on student learning: Towards a more competent theory for medical education', in *Innovation in medical education: An evaluation of its present status*, Z. Nooman, H. Schmidt, & E. Ezzat, eds. (New York, Springer Publishing Co, 1990)

Colliver, J. 'Effectiveness of Problem-based learning curricula: Research & theory', *Academic Medicine*, vol. 75, no. 3, (2000), pp. 259-266.

Delbecq, A. & Van den Ven, A. 'A group process model for problem identification and programme planning', *Journal of Applied Behavioral Sciences*, vol. 7, (1971), pp. 467-492.

Drummond, M. F. *Principles of economic appraisal in health care* (Oxford, Oxford University Press, 1980)

Engel, C. E. 'Not Just a Method but a Way of Learning', in *The challenge of problem based learning*, D. Boud & G. P. Felletti, eds. (London, Kogan Page, 1991) pp. 22-33

English National Board 1994, *Creating Lifelong Learners: Partnerships for Care*, English National Board, London.

Entwistle, N. 1992, *The impact of Teaching & learning Outcomes on higher Education*, Committee of Vice Chancellors and Principals, Sheffield.

Entwistle, N. J. & Ramsden, P. *Understanding student learning* (London, Croom Helm, 1983)

Evenson, D. & Hmelo, C. 'Introduction to Problem based learning: Gaining insights on learning interactions through multiple methods of enquiry', in *Problem based learning A research perspective on learning interactions*, D. Evenson & C. Hmelo, eds. (Mahwah, Lawrence Erlbaum, 2000)

Facione, P. A. *The California Critical Thinking Skills Test* (Millbrae, California Academic Press, 1990)

Feletti, G. P. 'Reliability and validity studies on Modified Essay Questions', *Journal of Medical Education*, vol. 55, (1980), pp. 933-941.

Felletti, G. P. & Smith, E. K. M. 'Modified essay questions: are they worth the effort?', *Medical Education*, vol. 20, no. 126, (1986), p. 132.

Fisher, A. & Scriven, M. *Critical Thinking: Its definition and assessment* (Point Reyes, Edgepress, 1997)

Fisher, M., King , J., & Tague, G. 'Development of a self-directed learning readiness scale for nursing education', *Nurse Education Today*, vol. 21, (2001), pp. 516-525.

Gijselaers, W. H. 'Connecting problem-based practices with educational theory', in *Bringing problem-based learning to higher education: Theory and Practice*, L. Wilkerson & W. Gijselaers, eds. (San Francisco, Jossey Bass, 1996) pp. 13-22

Goodstein, L., Cooke, P., & Goodstein, J. 'The team orientation and behaviour inventory (TOBI):', in *The 1983 annual for facilitators, trainers and consultants*,San Diego, University Associates, 1983) pp. 108-114

Howell Adams, M., Whitlow , J., Stover L, & Williams Johnson, K. 'Critical Thinking as an educational outcome: An evaluation of current tools of measurement', *Nurse Educator*, vol. 21, no. 3, (1996), pp. 23-32.

Hutchinson, L. 'Evaluating and researching the effectiveness of educational interventions', *British Medical Journal* no. 318, (1999), pp. 1267-1269.

Jaques, D. Small Group teaching: A SEDA Induction pack. 1990. Birmingham, SEDA.

Johnson, R. Marking essays: Guidance notes for tutors. 1993. Manchester, Clifton Press.

Knox, J. D. E. 'How to use modified essay questions', *Medical Teacher*, vol. 2, no. 1, (1980), pp. 20-24.

Kolb, D. A. *Experiential learning: Experience as the source of learning and development* (Englewood Cliffs, Prentice Hall Inc, 1984)

Long, M. & Johnson, T. 1997, *Influences on the Course Experience Questionnaire scales*, Australian Council for Educational Research, Sydney.

Louis, T., Lavori, P., Bailar, J., & Polansky, M. 'Cross-over and self-controlled designs in clinical research', *New England Journal of Medicine*, vol. 310, no. 1, (1984), pp. 24-31.

Margetson, D. 'The role of evaluation', in *Problem-based Learning in a Health Sciences Curriculum*, Alavi C, ed. (London, Routledge, 1995) pp. 141-161

Marks-Maran, D. & Gail Thomas, B. 'Assessment and evaluation in problem based learning', in *Problem-based learning in Nursing A new model for a new context*, S. Glen & K. Wilkie, eds. (Basingstoke, Macmillan Press, 2000) pp. 127-150

Marton, ., & Saljo R , ' Approaches to learning', in *The experience of learning*, F. Martin, D. Hounsell, & N. Entwistle, eds. (Edinburgh, Scottish Academic Press, 1984)

Maudsley, G. 'Do we all mean the same thing by "problem-based learning"?: A review of the concepts and formulation of the ground rules', *Academic Medicine*, vol. 74, no. 2, (1999), pp. 178-185.

McColl, E. & Jacoby A, T. L. S. J. B. C. S. N. e. a. Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. Health Technology Assessment 5[31]. 2001.

Mckeachie, W., Pintrich, P., Yi-Guand, L., & Smith, D. 1986, *Teaching & Learning in the college classroom: A review of the research literature*, National center for research to improve postsecondary teaching and learning, University of Michigan, Michigan.

Muller, S. 'Physicians for the 21st century : Report of the Project Panel of the General Professional Education of the Physician and College Preparation for Medicine', *Journal of Medical Education*, vol. 59 Part 2, (1984),

Myers Kelson, A. & Distlehorst, L. 'Groups in problem-based learning: Essential elements in theory and practice', in *Problem-Based Learning: A Research Perspective on Learning Interactions*, D. Evenson & C. Hmelo, eds. (Mahwah, Lawrence Erlbaum Associates, 2000) pp. 167-184

Norman GR, Powles, E., Powles, A., Rooney, P., Henry, N., & Dodd, P. 'Factors underlying performance on written tests of knowledge', *Medical Education*, vol. 21, (1987), pp. 297-304.

Norman, G. 'What should be assessed', in *The challenge of problem based learning*, D. Boud & G. Feletti, eds. (London, Kogan Page, 1991) pp. 254-259

Norman, G. & Schmidt, H. 'The psychological basis of Problem-based learning: A review of the evidence', *Academic Medicine*, vol. 67, no. 9, (1992), pp. 557-564.

Nowlen, P. *A new approach to continuing education for business and the professions: The performance model* (New York, MacMillan, 1988)

Oakley, A. *Experiments in knowing: gender and method in the social sciences* (Cambridge, Polity Press, 2000)

Painvin, C., Neufeld, V. R., & Norman, G. 1979, ' The triple jump exercise- a structured measure of problem-solving and self-directed learning', (Washington, 1979)

Pascarella, E. & Terenzeni, P. *How college affects students': Findings and insights from 20 years of research* (San Francisco, Jossey-Bass, 1991)

Patton, B., Giffin, K., & Patton, E. *Decision making group interaction*, 3rd edn, (New York, Harper Row., 1989)

Peters, A., Greenberger-Rosovsky, R., Crowder, C., Block, S., & Moore, G. 'Long-term outcomes of the new pathway program at harvard Medical School: A randomized controlled trial', *Academic Medicine*, vol. 75, no. 5, (2000), pp. 470-479.

Prescott, R., Gillespie, W., Counsell, C., Grant, A., Ross, S., Russell, I., & Kiauka, S. 'Factors that limit the number, quality and progress of randomised trials', in *Health Services Research Methods: A guide to best practice*, N. Black et al., eds. (London, BMJ Books, 1998) pp. 86-97

Quinn, R., Faerman, S., Thompson, M., & McGrath, M. *Becoming a master manager* (New York, John Wiley and Sons, 1990)

Ramsden, P. *Learning to teach in Higher Education* (London, Routledge, 1992)

Redding, N. 'Assessing the big outcomes', *Educational Leadership*, vol. May, (1992), pp. 49-53.

Rideout, E. 'Evaluating student learning', in *Transforming Nursing Education through Problem-based learning*, Rideout E, ed. (Sudbury, Jones & Bartlett, 2001) pp. 215-239

Rideout, E. & Carpio, B. 'The problem based learning model of nursing education', in *Transforming Nursing Education through problem-based learning*, Rideout E, ed. (Sudbury, Jones & Bartlett, 2001) pp. 21-45

Robson, C. *Real World research: A resource for social scientists and practitioner researchers* (Blackwell, Oxford, 1993)

Rolfe, I. & Murphy L., M. J. ' The interaction between clinical reasoning, group process and problem based learning assessment at the Newcastle Medical School', in *Research & development in Problem Based Learning: Proceedings of the 2nd bi-annual conference of Australian Problem Based Learning Network*, M. Ostwald & A. Kinmouth, eds. (Sydney, Charles Sturt University Press, 1994)

Saunders, N., Engel, C. E., & Feletti, G. 'Clinical Supervisors Report', *Medical Teacher*, vol. 4, no. 4, (1982), pp. 151-154.

Savin-Baden, M. *Problem-based learning in Higher Education: Untold Stories* (Buckingham, Society for Research in Higher Education and Open University Press, 2000)

Schmidt, H. G. 'Foundations of problem based learning; some explanatory notes', *Medical Education*, vol. 27, no. 422, (1993), p. 432.

Selltiz, C., Wrightsman, L., & Cook, S. *Research methods in social relations* (New York, Holt, Rinehart and Winston, 1976)

Senn, S. *Cross over trials in clinical research* (John Wiley, London, 1993)

Snellen-Balendeng, H. & Schmidt, H. 'Perceived workload as an indicator of the quality of education: The case of Maastricht;', in *Innovation in medical education: An evaluation of its present status*, Z. Nooman, H. Schmidt, & E. Ezzat, eds. (New York, Springer Publishing Co., 1990) pp. 390-403

Swanson D , 'A measurement framework for performance based tests', in *Further developments in assessing clinical competence*, I. Harvey & R. Harden, eds. (Montreal, Can-Heal, 1987) pp. 13-45

Swanson, D., Case, S., & Van der Vleuten CPM , ' Strategies for student assessment', in *The Challenge of Problem-based Learning*, D. Boud & G. Feletti, eds. (London, Kogan page, 1991) pp. 260-274

Tait, H. & Entwistle, N. 'Identifying students at risk through ineffective study strategies', *Higher Education*, vol. 31, (1996), pp. 97-116.

Tate, R. 'Experimental design', in *Encyclopaedia of education research*, H. Mitzel & W. Rabinowitz, eds. (London, Collier Macmillan, 1982) pp. 553-561

Thomas, R. E. 'Problem-based learning: measurable outcomes', *Medical Education*, vol. 31, (1997), pp. 320-329.

Tompkins, C. 'Nursing education for the 21st Century', in *Transforming Nursing Education through problem-based learning*, Rideout E, ed. (Sudbury, Jones & Bartlett, 2001) pp. 1-21

Torgerson, C. & Torgerson, D. 'The need for randomised controlled trials in educational research', *British Journal of Educational Studies*, vol. 49, no. 3, (2001), pp. 316-318.

Van Der Vleuten, C. P. M. 'The assessment of professional competence: Developments, research and practical implications', *Advances in Health Sciences Education*, vol. 1, no. 41, (1996), p. 67.

Verma M, C. J. & Singh, T. 'Reliability of essay type questions- effect of structuring', *Assessment in Education.*, vol. 4, no. 2, (1997), pp. 265-270.

Vernon D.T, A. & Blake, R. L. 'Does Problem-based Learning Work? A Meta-analysis of Evaluative Research', *Academic Medicine*, vol. 69, no. 7, (1993), pp. 550-563.

Walton J, C. D. C. N., Walton, J., Clark, D., & Click, N. 'An outcomes assessment of a hybrid-PBL course in treatment planning', *Journal of Dental Education* (1997), pp. 361-367.

Walton, H. J. & Matthews, M. B. 'Essentials of Problem Based Learning', *Medical Education*, vol. 23, (1989), pp. 542-558.

Wilkes, M. & Bligh, J. 'Evaluating educational interventions', *British Medical Journal*, vol. 318, (1999), pp. 1260-1272.

Wilkie, K. 'The nature of Problem-based learning', in *Problem based learning in nursing*, S. Glen & K. Wilkie, eds. (Basingstoke, Macmillan Press, 2000) pp. 11-34

Woods, D. R. *Problem-based learning: helping your students gain the most from PBL* (Hamilton CA: DR Woods., 1995)

Woodward, C. 'What can we learn from programme evaluation studies in medical education', in *The challenge of problem based learning*, 2nd edn, D. Böud & G. Feletti, eds. (London, Kogan page, 1997) pp. 294-307

Woodward, C. & Ferrier, B. 'Perspectives of graduates two or five years after graduation from a three year medical school', *Journal of Medical Education*, vol. 57, (1982), pp. 294-302.

World Health Organization 1993, *Increasing the relevance of Education for Health professionals: Report of a WHO study group on Problem Solving Education for Health professionals*, WHO, Geneva.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM034861

I. DOCUMENT IDENTIFICATION:

Title:	Evaluating educational impact: The approach followed by The Project on the Effectiveness of Problem Based Learning (PEPBL)
Author(s):	Mark Newman, Kate Ambrose, Trevor Corner, Jeff Evans, Phyl Morris-Vincent, Sheila Quinn, John Stephenson, Lesley Vernon
Corporate Source:	Middlesex University (UK)
Publication Date:	~~2#~~ '2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature:	Printed Name/Position/Title: Mr Mark Newman	
Organization/Address: Middlesex University, Furnival Building, 2-10 Highgate Hill, London N.19 5LW United Kingdom	Telephone: 0044 208 411 6702	FAX: 0044 208 411 6299
	E-Mail Address: M.newman@mdx.ac.uk	Date: 24/4/2003

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions